Palaeogeography, Palaeoclimatology, Palaeoecology 506 (2018) 22–29

Contents lists available at ScienceDirect

Palaeogeography, Palaeoclimatology, Palaeoecology

journal homepage: www.elsevier.com/locate/palaeo

Chronological uncertainty severely complicates the identification of cyclical processes in radiocarbon-dated time-series

W. Christopher Carleton[a], David Campbell[b], Mark Collard[a,*]

[a] Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada
[b] Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

ARTICLE INFO

Keywords:
Palaeoenvironmental cycles
Sociopolitical cycles
Time-series analysis
Radiocarbon dating
Calibration
Temporal uncertainty

ABSTRACT

Cycles are widely considered to be an important feature of environmental and human history over the last 50,000 years. However, there is an overlooked problem in the investigation of cyclicity in this time period—the standard statistical methods for identifying cycles assume that observations are precisely dated, but the main relevant dating technique, radiocarbon dating, often yields dates with large and highly irregular uncertainties. Here, we present the results of a massive simulation study that explored the impact of radiocarbon dating uncertainty on our ability to identify cycles in time-series. Our results suggest there is indeed a problem. We found that, at best, we could correctly identify known cycles only 42% of the time and that the false-positive rate was as high as 90%. This indicates that an individual analysis of a single time-series is very likely to return false-positive cycles. One implication of this is that many of the environmental and sociopolitical cycles that have been identified may not be real. Consequently, a program of reassessment is needed.

1. Introduction

Cyclicity is an important concept in the study of the last 50,000 years. Numerous cyclical phenomena have been identified in the recent history of Earth's climate, including Dansgaard-Oeschger Events, Heinrich Stadials, and Bond Cycles (Bianchi and McCave, 1999; Bond, 1997; deMenocal, 2000; Desprat et al., 2003; Hu, 2003; Langdon et al., 2003; Moreno et al., 2005; Sorrel et al., 2012). Cycles are also widely thought to be an important feature of human history (Collingwood, 1927; Gavrilets et al., 2010; Gronenborn et al., 2014; Redman and Kinzig, 2003; Rosen and Rivera-Collazo, 2012; Sandoval, 1998; Thompson and Turck, 2009; Turchin and Nefedov, 2009; Zhang et al., 2006; Zimmermann, 2012). For example, both the Classic Maya and the ancient Mesopotamians have often been said to have experienced sociopolitical cycles (Hodell et al., 2005; Marcus, 1992; Masson, 2012; Ur, 2010).

Identifying environmental and sociopolitical cycles normally involves time-series analysis. A time-series is simply a sequential set of observations where the chronological arrangement of the observations is important (Koopmans, 1974). Most established statistical methods of searching for cycles in time-series draw on Fourier Theory, which holds that any continuous function of real numbers can be decomposed into sums of sine and cosine waves (Koopmans, 1974). In practice, most established time-series methods involve fitting a set of waves with different frequencies to a time-series either directly with regression or via the Fast Fourier Transform. The frequencies that account best for the variance in a given time-series are deemed to represent important cycles.

It is usually assumed that standard time-series methods can be used straightforwardly to identify cyclical patterns in palaeoenvironmental and archaeological time-series (e.g., Bianchi and McCave, 1999; Bond, 1997; deMenocal, 2000; Desprat et al., 2003; Hodell et al., 2005). However, this assumption is problematic in relation to time-series from the last 50,000 years. The problem is that most such time-series are dated with the radiocarbon method and radiocarbon dating often yields large and highly irregular uncertainties (Telford et al., 2004)(e.g., Fig. 1). This is because radiocarbon dates have to be calibrated to account for the through-time variation in the atmospheric $^{14}C/^{12}C$ ratio (Buck et al., 1996), and the calibration curve contains fluctuations and plateaus that cause a radiocarbon year to correspond to multiple calendar years with similar probabilities. One of the effects of multimodal, highly skewed uncertainties is that the amount of time between observations in a given time-series will be uncertain. This uncertainty means multiple waveforms with different frequencies can fit equally well, making it difficult to securely identify cycles.

To date, the impact of impact of calibrated radiocarbon dates on time-series analysis has not been investigated. The significance of chronological uncertainty for time-series analysis has been recognized

* Corresponding author.
E-mail address: mcollard@sfu.ca (M. Collard).

https://doi.org/10.1016/j.palaeo.2018.06.002
Received 7 March 2018; Received in revised form 29 May 2018; Accepted 1 June 2018
Available online 03 June 2018

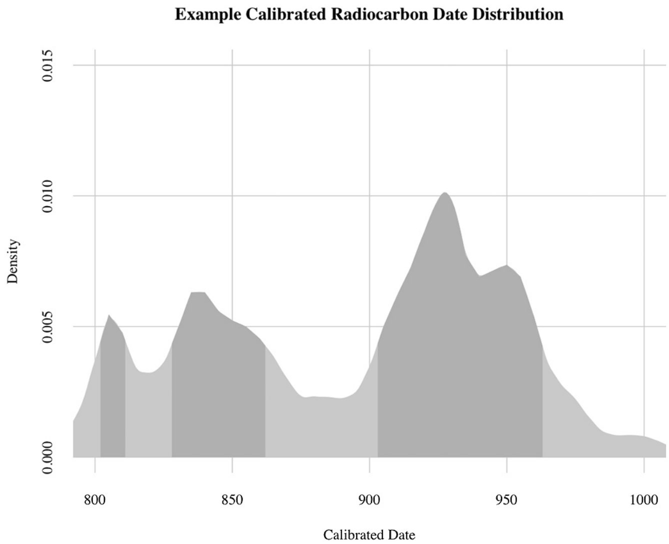

Fig. 1. Examples of calibrated radiocarbon date distribution showing the highly irregular nature of the errors—darker regions correspond approximately to the 68% highest density region (i.e., there is a 68% chance that one of the dates in the shaded areas corresponds to the true calendar date).

by some researchers, but the specific impact of radiocarbon dating uncertainty has not been examined. For example, in 2009 Mudelsee and colleagues updated their seminal computer software for identifying cycles in palaeoclimate records—REDFIT—to account for general chronological uncertainty, but they did not explore the impact of radiocarbon dating uncertainty on the new software. A few years later, Mudelsee (2014) showed that chronological uncertainty in general can lead to incorrect estimates of confidence intervals in a wide range of time-series methods, including common frequency-based ones, but did not discuss the specific problems caused by radiocarbon dates. In a similar vein, Martinez et al. (2016) looked at the impact of stratigraphic uncertainty on frequency-based time-series analysis but did not examine the problems introduced by the use of radiocarbon dating to generate age-depth models. More recently, Franke et al. (2018) introduced a Bayesian frequency-based time-series method in an attempt to account for chronological uncertainty, but these authors also did not explore the impact of the multimodality associated with calibrated radiocarbon dates on cycle identification.

With the foregoing in mind, we carried out a study in which we used massive simulation experiments to assess the error rates that can be expected when searching for cycles in frequency-based time-series dated with calibrated radiocarbon dates (i.e. a statistical power analysis). The experiments involved the creation of synthetic time-series with the same basic features as the main type of time-series used in palaeoenvironmental and archaeological research—sets of ratio observations dated with a sample of radiocarbon dates. The synthetic time-series can be thought of as, say, artificial temperature proxy records or artificial estimates of population density derived from room counts. They comprised a simple sine wave and autocorrelated noise with 300 observations spanning 1000 years. Each time-series contained five, ten, or 40 complete cycles—i.e., oscillations—of the synthetic sine wave, which determined the frequency range of the wave. The number of oscillations is important because it affects the statistical significance of the results. More oscillations in a given time-series means that the wave should be easier to identify given a sufficient sampling rate (up to the Nyquist limit [Bloomberg, 2000]) making the evidence for a given cycle more compelling. Five, ten, and 40 oscillations were selected to reflect typical radiocarbon-dated time-series used in palaeoenvironmental and archaeological research (Table S1). In the experiments, we sought to recover the frequency of the known sine waves given different degrees of uncertainty.

Table 1
Simulation parameters.

Variable name	Value(s)
Time-series length	300 observations
Timespan	1000 years
Autocorrelation coefficient	0.6
Complete oscillations	5, 10, 40
Frequencies	0.005, 0.01, 0.04 (oscillations per year)
Signal-to-noise ratio	1, 10, 100
Number of dates	5, 10, 15
Location on the calibration curve	12,000–13,000 cal. yr B.P.; 14,000–15,000 cal. yr B.P.

2. Materials and methods

We conducted the experiments in the R statistical computing environment (R Core Team, 2017). The experiments explored how several variables affect statistical power—i.e., the ability to identify the known sine waves in our synthetic time-series (see Supplementary Information for R scripts). Each experiment involved four variables: 1) the amount of noise; 2) the number of oscillations of the wave; 3) the number of simulated radiocarbon dates; and 4) whether the dates for the time-series were calibrated with a high-slope section of the radiocarbon calibration curve or a low-slope section. These variables were chosen because statistical time-series theory and research on radiocarbon-dated age-depth models suggest they may affect the degree of chronological uncertainty and the clarity of the cyclical pattern (Bloomfield, 2000; Telford et al., 2004; Franke et al. 2018). Details of the parameters and their values are given in Table 1.

The simulation proceeded in several steps. First, we created 1000 synthetic time-series for each combination of parameters (Fig. 2). For the purposes of this study, the specific values employed are not as important as the relationships among the variables. The number of cycles per unit time illustrates why this is the case. As long as the sampling density is constant, 40 oscillations in 1000 years is the same as 400 oscillations in 10,000 years or four in 100 years. Even though the specific values employed are not as important as the inter-variable relationships, we selected parameter settings that are in line with data

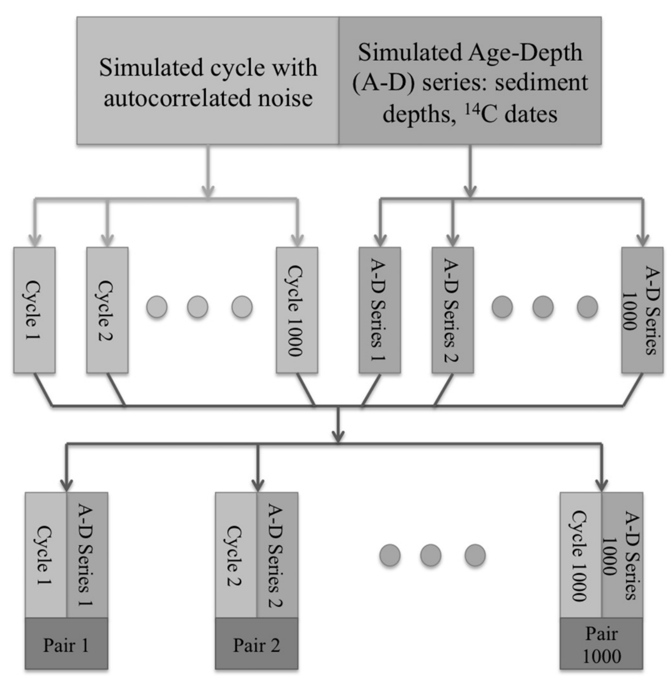

Fig. 2. Flowchart of the simulation process: pairing 1000 samples of a given time-series cycle to simulated age-depth data.

and problems relevant to palaeoenvironmental and archaeological research (e.g., Carleton et al., 2014). We also ensured that the parameter settings covered the range of variation we found among a selection of palaeoenvironmental time-series available on the website of the National Oceanic and Atmospheric Administration (www.noaa.gov; see Table S1).

Each time-series comprised a simple sine wave and additive auto-correlated or "red" noise. The red noise was created with an R function called *arima.sim*() that generates realizations of an auto-regressive process. The standard deviation of the noise process determined the signal-to-noise ratio of the time-series.

Along with the simulated observations, we created a set of artificial radiocarbon dates for each experiment. The dates were evenly spaced along the calibration curve within one of two fixed segments—i.e., 12,000–13,000 cal yr B.P. or 14,000–15,000 cal yr B.P. These segments were chosen because the latter had roughly twice the slope of the former (see Fig. 4), which allowed us to explore the effect of the calibration curve's slope on the results. We set the error of the simulated radiocarbon dates to ± 50 years, which is a common magnitude of uncertainty returned by radiocarbon dating labs. We used the same error for all the dates in order to isolate the uncertainties introduced by the calibration process.

In the second step of the simulation, we randomly sampled the calibrated date distributions of each time-series 2000 times (Fig. 3). This yielded a total of 2,000,000 time-series per experiment. Any set of simulated dates that violated the simulated stratigraphic relationships among the dates was discarded. This makes the procedure we employed equivalent to Bayesian calibration (Buck et al., 1996). Each sample of dates was interpolated with a monotonic spline to provide a date for each of the observations in a given time-series. We opted to use splines because recent research indicates that they are the most stable, robust type of age-depth model when radiocarbon dating is used (Telford et al., 2004).

In the simulation's third step, we analyzed the time-series using

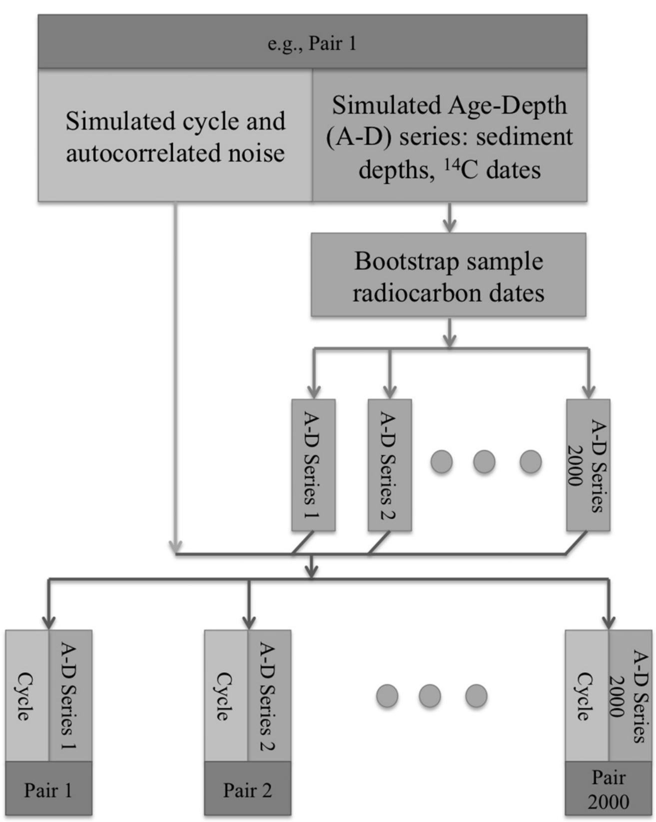

Fig. 3. Flowchart of the simulation process: the radiocarbon date bootstrap.

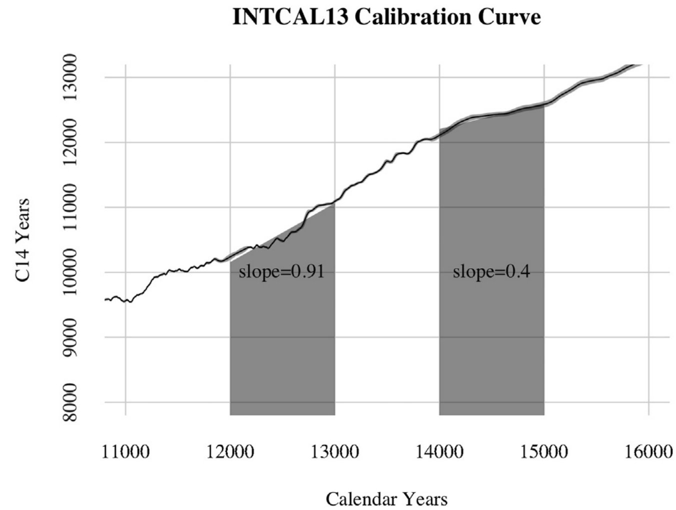

Fig. 4. Calibration curve regions used in the simulation.

Least-Squares Spectral Analysis (LSSA), which is an established method for finding cycles in unevenly spaced time-series (Lomb, 1976; Vaníček, 1971). Equivalent to the well-known Lomb-Scargle method (Bretthorst, 2003; Lomb, 1976), LSSA potentially has broad applicability to palaeoenvironmental and archaeological time-series because of its ability to cope with unevenly spaced data. Like other methods for finding cycles, LSSA produces a frequency spectrum—i.e., a plot with peaks identifying the frequency of waveforms that might be present in a given time-series. In contrast to most other spectra, however, a Least-Squares spectrum indicates the correlation of the fit between a wave at a given frequency and the time-series under analysis. We compared the LSSA spectra yielded by the synthetic time-series to an obvious benchmark for statistical significance—a null frequency spectrum that reflects our expectations about how a random time-series with no cycles would appear. Importantly, the null spectrum had to account for auto-correlation because environmental and sociopolitical processes are typically temporally autocorrelated, resulting in a red noise background (Schulz and Mudelsee, 2002). We produced the null spectra by analyzing 1000 additional simulated time-series for each experiment. These time-series contained no sine waves, only red noise created in the same way as before. Using LSSA, we estimated a spectrum for each of the 1000 red-noise-only time-series and then extracted the 95% quantile of each frequency in the spectra. We identified peaks in the frequency spectrum of a given time-series as statistically significant only if they exceeded this 95% level, which is equivalent to the commonly used p-value of 0.05. While a more stringent criterion would arguably be better, the 0.05 threshold is still the most commonly used benchmark for statistical significance, making it appropriate for our study. We allowed for a generous window of error by including any peaks in the spectrum within ± 20% of the known frequency of the sine wave used to create time-series. Peaks that fell within the window of error were considered "hits"—i.e., true-positives.

Lastly, we collated the hits and plotted them, creating hit-rate distributions. The distributions illustrate the number of times a given frequency was found to be significant for a given combination of simulation parameters. Permuting all possible values for the four free parameters resulted in 54 experiments. This required substantial high-performance computing power involving more than a hundred processors and several months of computational time.

3. Results and discussion

The experiments produced several results (Fig. 5 and S1-S5). Some of them were unsurprising. For example, we found that increasing the noise in the time-series led to more false-positive findings, and that

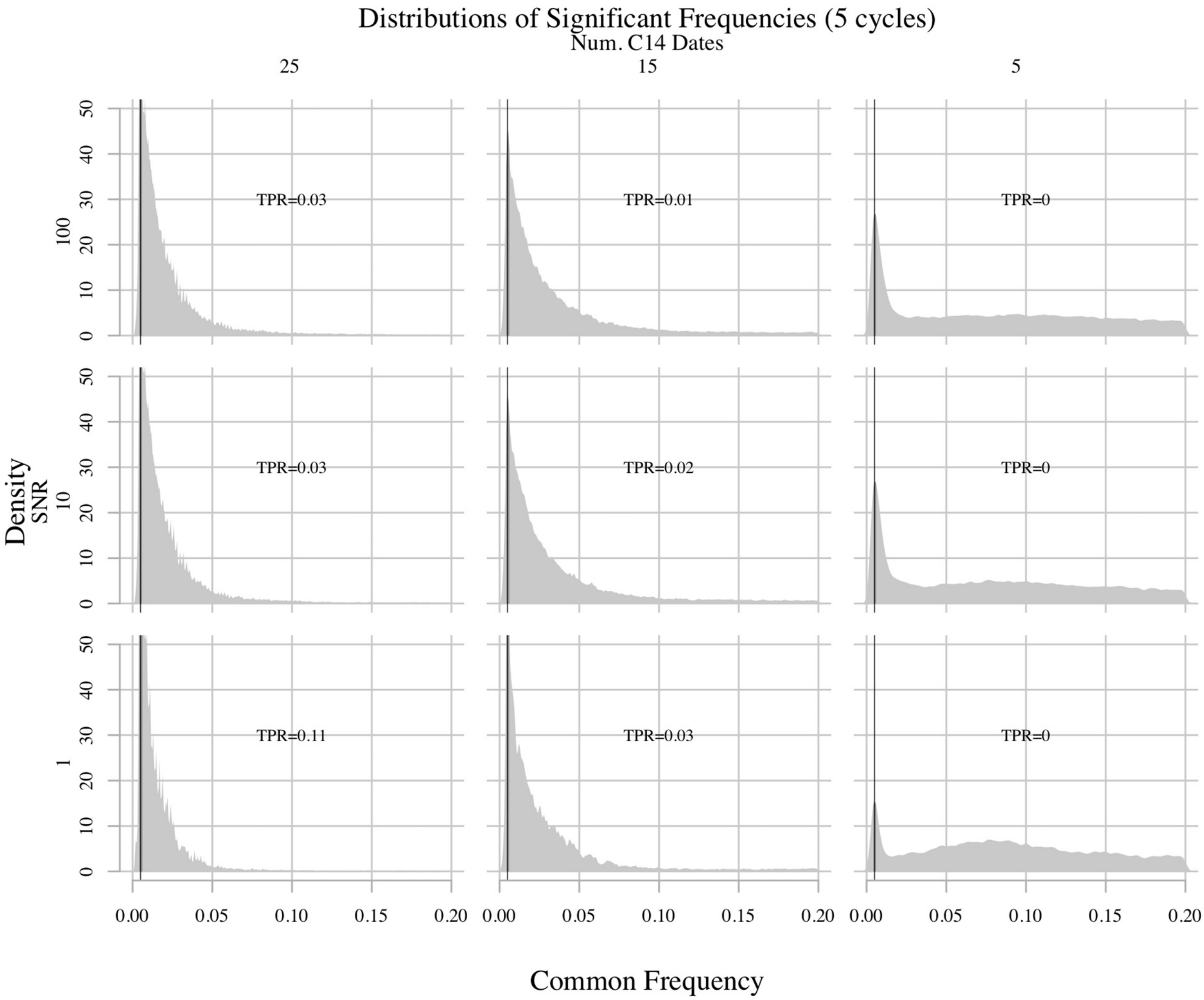

Fig. 5. Significant frequencies identified in experiments involving 5 oscillations between 12,000 and 13,000 cal yr B.P. The vertical black lines denote the target frequency. Signal-to-noise ratio varies among the panels vertically while the number of radiocarbon dates varies horizontally. The rest of the figures summarizing our results are in the Supplementary Information.

more oscillations were easier to detect than fewer oscillations. Similarly, we found that increasing the number of radiocarbon dates improved the hit rates, and that experiments involving the low-slope portion of the calibration curve produced greater numbers of false-positives. However, the hit rates and false-positive rates were surprising. Specifically, they were surprisingly bad. Around two-thirds of the experiments yielded hit rates of less than 10%, while the false-positive rates reached 90% in two-thirds of the experiments despite specifying the Type I error rate with a *p*-value significance threshold of 0.05. This means that it is highly likely that an individual analysis of a single time-series would yield false-positive cycles. Together, the hit rates and false-positive rates suggest that chronological uncertainty substantially undermines our ability to securely identify cycles in radiocarbon-dated time-series when point estimates of the dates are used to build age-depth models.

It is possible that the situation may be even worse than our experiments suggest. First, as we noted earlier, we used a wide window of error for counting a true-positive finding. The error window meant that, for instance, the target period for a 25-year cycle was 25 ± 5 years. A longer cycle of 200 years would have a window of 200 ± 40 years, leaving a lot of room for different candidate cycles. Obviously a narrower window of error would have made it harder to correctly identify cycles.

Second, many real palaeoenvironmental and archaeological datasets are likely much more complicated than a single waveform plus red-noise. Real environmental and sociopolitical processes could conceivably involve multiple cycles with various wavelengths. In such cases, distinguishing between multiple cycles would be harder than identifying a single cycle, especially if the overlapping cycles have similar periods. Real environmental and sociopolitical processes might also contain acyclic trends or rapid fluctuations—i.e., "jumps" that occur over short intervals. Distinguishing cycles from these acyclic trends and fluctuations would be more challenging than identifying cycles in simple time-series like those in our simulation experiments.

Lastly, the highest hit rates we found were for experiments involving low noise levels. In the experiment with a 42% hit rate, the time-series was almost a pure sine wave. Most real-world palaeoenvironmental datasets are substantially noisier because of chaotic climate fluctuations and instrumentation error. The same holds for most archaeological datasets. So, for real-world applications we expect the false-positive rate to go up and the true-positive rate to go down. Thus, there is another reason to think that securely identifying unknown cycles might be even less likely than our study suggests.

It seems like increasing the number of dates used to create age-depth models should solve the problem. In theory, more dates should mean less temporal uncertainty about the times associated with individual observations, and thus the time between observations in a given series should be more certain, reducing the number of potential waveforms that could fit the series. However, our results suggest the situation is more complicated than that. While we found that increasing

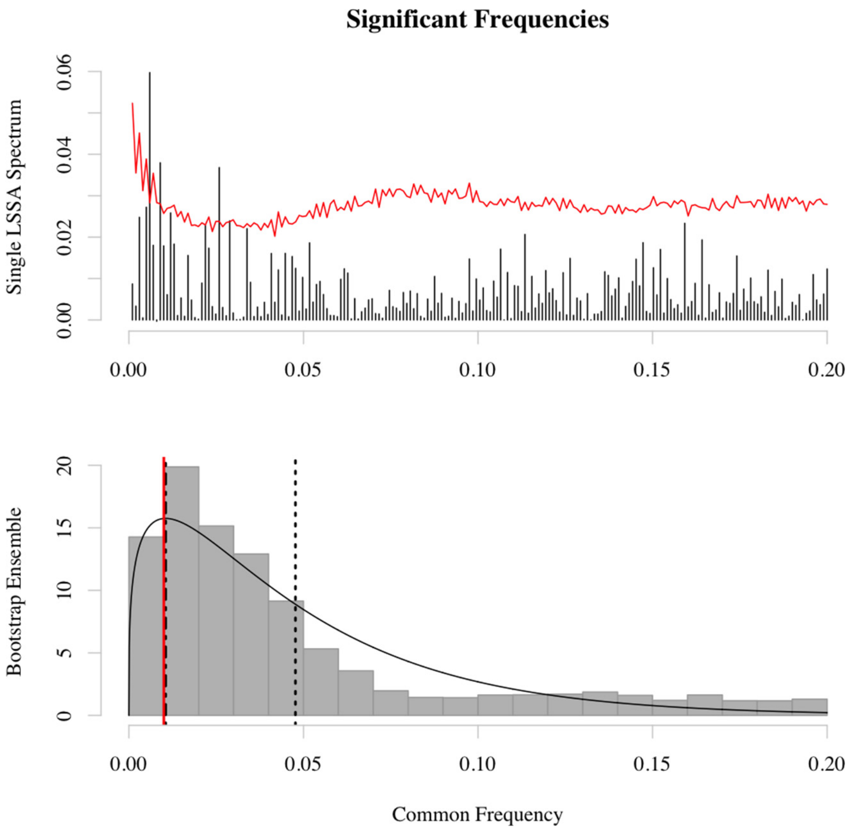

Fig. 6. Example of significant frequencies from a single LSSA spectrum (top panel with the 95% confidence level shown as a red line) and the distribution of significant frequencies found during a radiocarbon-date bootstrap (bottom panel). In the bottom panel, a Gamma distribution is drawn over the histogram. The distribution parameters were estimated with maximum likelihood methods from the pool of significant frequencies. The vertical dash-dot line indicates the mode of the Gamma distribution, which corresponds to the most commonly occurring significant frequency; the vertical dashed line indicates the mean of the Gamma distribution; and the red vertical line indicates the target frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the number of dates improved the hit rates, the relationship between numbers of dates and hit rates was weak. Increasing from 10 to 15 dates, for example, had only a small impact on hit rates across experiments—the impact was, in fact, barely noticeable (see Fig. 5 and Figs. S1–S5).

We suspect that the size of the improvement in hit rate might be negligible because of the highly irregular nature of calibrated radiocarbon dating uncertainty. Increases in date density might not produce significant reductions in age-depth model uncertainty because typical calibrated radiocarbon date distributions contain multiple modes, long

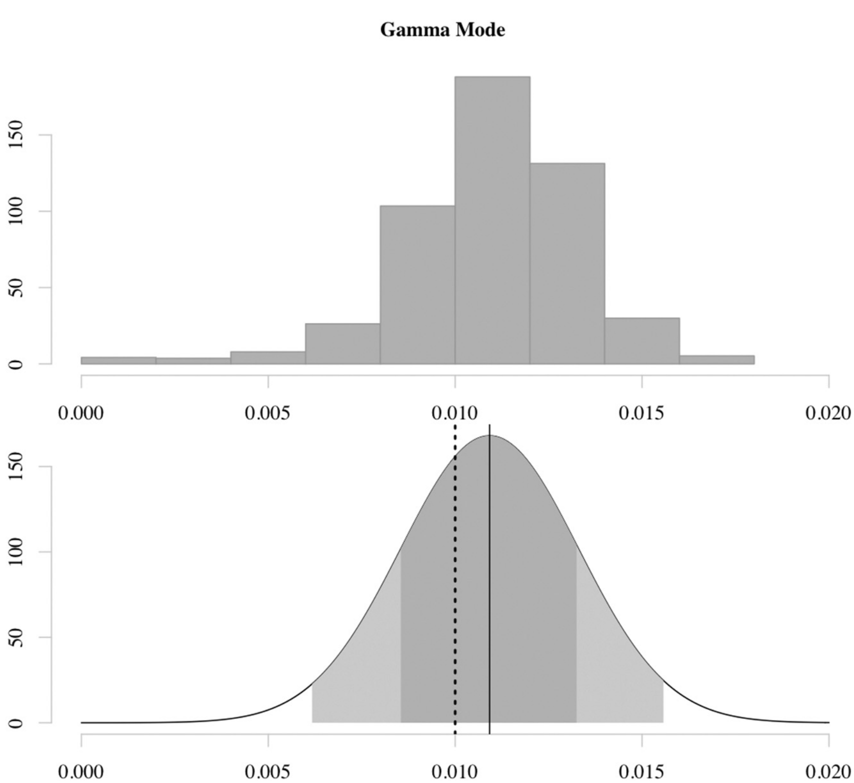

Fig. 7. The distribution of modes for all 1000 simulated time-series. The top panel is the empirical histogram. The bottom panel shows a normal approximation of the distribution of modes with a vertical black line showing the mean and the vertical dashed line showing the target frequency. This distribution was estimated with maximum likelihood methods, but it is clear that the empirical data are skewed, which is why we opted not to use the approximation in the bottom panel. It is shown here because it illustrates the proximity between the target frequency and the mean of the modes.

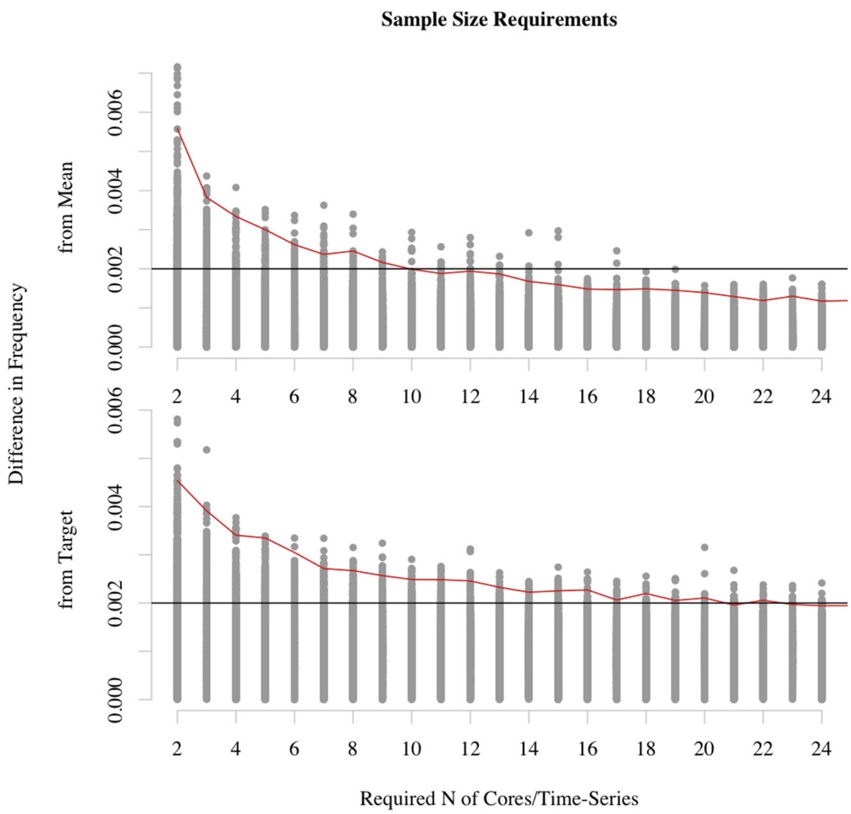

Fig. 8. Empirically derived sample sizes (x-axis) required for different margins of error (y-axis). These plots show the margin of error (difference) between the mean of the modes from a given sample and the mean of all modes (top panel) or the target frequency (bottom panel). Each dot represents the difference from a single sample of a given sample size, indicated on the x-axis. The red line indicates the 95% confidence level and the black horizontal line indicates approximately a 20% margin of error for a cycle with a frequency of 0.01, or a period of 100 years. Both panels clearly show that increasing the sample size (number of time-series) leads to lower differences between the sample mean—of modal significant frequencies—and the population estimates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

plateaus, and marked skews. This irregular uncertainty leads to many equally likely age-depth models with highly variable features over short and long time-spans even with many dates acting as chronological anchors. This type of variability can be expected to hinder statistical convergence toward a single, true age-depth model even as the dating density is increased—i.e., the chronological evidence will not necessarily clearly indicate a single true model as we increase the number of (highly irregular) chronological anchors. If this is the case, it may take a very large number of additional radiocarbon dates before a noticeable improvement in hit rates is achieved. This may not be the case with dating methods that yield more regular uncertainties. While further research is certainly required, our present findings indicate that for moderate numbers of dates increasing the number of radiocarbon dates is not sufficient to improve cycle identification.

Thankfully, our findings also suggest a way to at least partially address the problem in another way. Recall that the hit-rate distributions represent pooled results from 2,000,000 individual frequency analyses, each of which involved a single time-series with a set of dates randomly sampled from the relevant calibrated date distributions. These hit rate distributions (e.g., Fig. 5) are unimodal with peaks close to the target frequencies. Significant frequencies around these peaks occur more often than other frequencies, and the probability of a given frequency appearing significant in our simulation declines with distance from these peaks. This means that, as the number of time-series increases, a pooled estimate should converge to the modal frequency, which is generally close to the target frequency, as we noted earlier. Thus, while a frequency analysis of one time-series is likely to yield significant numbers of false-positives and false-negatives, combining results from multiple time-series should lead to a better estimate.

To evaluate this possibility, we ran a further experiment. The new experiment involved a cycle with a frequency of 0.01 and 15 radiocarbon dates. Instead of combining all of the results as we did before, we looked at the distributions of significant frequencies for each of the 1000 simulated time-series. The distributions comprised the significant frequencies identified during the radiocarbon-date bootstrap simulations. We found that these distributions were also unimodal with peaks reasonably close to the target frequency (Fig. 6). To determine how many time-series would be required to estimate the target frequency, we fit Gamma distributions to the significant frequencies identified with the radiocarbon-date bootstrap. We then calculated the mode of the Gamma distribution for each one, creating a list of 1000 modes. Each mode corresponded to the peak of the relevant distribution of significant frequencies (Figs. 6 and 7). The distribution of modes was close to Normal (Fig. 7), but did not pass the standard tests—it was left-skewed. So, instead of using the normal approximations for sample size given a margin of error, we simulated the sampling process from the list of modes. We randomly sampled the list of modes with increasing sample sizes ranging from two to 20. Each time we increased the sample size, we randomly sampled the list of modes without replacement 1000 times. Then, we calculated the difference between the mean of each sample and two values: 1) the mean of all modes, and 2) the target frequency (Fig. 8).

The results of the new experiment were encouraging. They indicated that increasing the number of time-series does indeed cause the estimate of the target frequency to converge. As a corollary, the estimate also closes in on the target frequency (Fig. 8, bottom panel). More importantly, we were able to determine that around 10–12 time-series would be required to estimate the population mean of modal frequencies with approximately a 20% margin of error at the 95% confidence level. Since our experiment involved a cycle with a frequency of 0.01, using 10 time-series to identify it would result in an estimate of the target frequency that was between 0.008 and 0.012 roughly 95% of the time. It would be possible to decrease the margin of error by increasing the number of time-series, but as Fig. 8 shows, there would be diminishing returns from doing so as the number of time-series required increases rapidly.

Our estimate of the required number of time-series makes two crucial assumptions. One is that the time-series in question are independently dated. Each of the 1000 time-series we simulated was created as if it were independent. This is similar to measuring time-

series from 20 different sediment cores. The other assumption is that there is only one significant cycle to find. Multiple cycles might produce multiple modes and discerning these modes from one another and from false-positives could be difficult. This is something that should be explored in future simulation work. Still, our results at least provide an initial benchmark and suggest a practicable strategy for partially overcoming the problems we identified with radiocarbon-dated time-series.

With regard to other future research directions, as we explained earlier, we created the synthetic time-series in such a way that they had characteristics that are similar to those of several high-profile published time-series (Table S1). Given this, an obvious implication of our study is that a number of widely discussed cycles could be spurious. Our findings suggest that the Type I error rates for statistical tests for significant frequencies are much larger than the specified 0.05 when dealing with realistic palaeoenvironmental and archaeological data. This implies that our understanding of the past climate system may not be as good as we have come to believe. The same holds for our understanding of sociopolitical cycles. Consequently, there would appear to be a need to re-analyze any radiocarbon-dated time-series that has been argued to support a putative natural or cultural cycle using a simulation approach capable of taking into account the impact of chronological uncertainty on cycle identification.

4. Conclusions

The study reported here investigated the impact of chronological uncertainty on our ability to the identify cycles in palaeoenvironmental and archaeological datasets that are dated with radiocarbon dates. Identifying cycles involves the use of time-series methods. These methods assume that the target dataset is precisely dated, but this assumption is often violated by radiocarbon-dated palaeoenvironmental and archaeological datasets. This is because the calibration process that is used to convert radiocarbon dates into calendar dates frequently results in highly irregular uncertainties. To assess the scale of the problem, we conducted a set of massive simulation experiments that involved estimating the statistical power of an established method for identifying cycles in the face of calibrated radiocarbon date uncertainty.

The experiments indicate that there is indeed cause for concern. Around two-thirds of the experiments yielded hit rates of less than 10%, while the false-positive rates reached 90% in two-thirds of the experiments despite specifying the Type I error rate with a *p*-value significance threshold of 0.05. Our results indicate that an individual analysis of a single time-series would very likely yield false-positive cycles. Together, the low hit rates and high false-positive rates suggest that chronological uncertainty substantially undermines our ability to securely identify cycles in radiocarbon-dated time-series when point estimates of the dates are used to build age-depth models. The obvious corollary of this is that many of the environmental and sociopolitical cycles that are thought to be features of the last 50,000 years could well be spurious.

The experiments allow us to make two recommendations for improving the chances of finding cycles in future research. One is that researchers should use multiple time-series. Our analyses demonstrated that increasing the number of time-series representing a given historical or climatic process improves our ability to correctly identify cycles, at least under certain conditions. As the number of time-series increases, a given analysis can be expected to converge to the modal frequency, which our simulation suggests would likely be close to the true frequency of a given cycle if it exists.

The other recommendation is to spread chronometric resources across multiple time-series rather than investing heavily in dating a single core. Specifically, we found that using more than 15 evenly spaced dates per 1000-years made little difference to the results, which could be used as a practicable benchmark for dating density. Utilizing more dates did not improve our ability to find cycles enough to compensate for the other common sources of uncertainty. Thus, if a project has limited funding, it is probably better to gather multiple dated time-series than a single time-series with a large number of dates.

Acknowledgments and notes

We thank Paul Hesse, Thomas Algeo, and several anonymous reviewers for their assistance. Their constructive criticisms and advice greatly improved our paper. We received financial support from the Social Sciences and Humanities Research Council of Canada (767-2011-0159 and 895-2011-1009), the Natural Sciences and Engineering Research Council of Canada (RGPIN04040-2014), the Canada Research Chairs Program (228117 and 231256), the Canada Foundation for Innovation (203808), the British Columbia Knowledge Development Fund (862-804231), and Simon Fraser University (14518).

We have no conflicts of interest to declare.

Additional information about the simulation experiments can be found in the Supplementary Information for this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.palaeo.2018.06.002.

References

Bianchi, G.G., McCave, I.N., 1999. Holocene periodicity in North Atlantic climate and deep-ocean flow south of Iceland. Nature 397, 515–517. http://dx.doi.org/10.1038/17362.

Bloomfield, P., 2000. Fourier Analysis of Time Series: An Introduction. Wiley, New York.

Bond, G., 1997. A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. Science 278, 1257–1266. http://dx.doi.org/10.1126/science.278.5341.1257.

Bretthorst, G., 2003. Frequency estimation and generalized Lomb-Scargle periodograms. In: Feigelson, E.D., Babu, G.J. (Eds.), Statistical Challenges in Astronomy. Springer, New York, pp. 1–18.

Buck, C.E., Cavanagh, W.G., Litton, C.D., 1996. Bayesian Approach to Interpreting Archaeological Data. Wiley, Chichester, UK.

Carleton, W.C., Campbell, D., Collard, M., 2014. A reassessment of the impact of drought cycles on the classic Maya. Quat. Sci. Rev. 105, 151–161. http://dx.doi.org/10.1016/j.quascirev.2014.09.032.

Collingwood, R.G., 1927. Oswald Spengler and the theory of historical cycles. Antiquity 1, 311–325. http://dx.doi.org/10.1017/S0003598X00000685.

Desprat, S., Sánchez Goni, M.F., Loutre, M.F., 2003. Revealing climatic variability of the last three millennia in northwestern Iberia using pollen influx data. Earth Planet. Sci. Lett. 213, 63–78. http://dx.doi.org/10.1016/S0012-821X(03)00292-9.

Gavrilets, S., Anderson, D.G., Turchin, P., 2010. Cycling in the complexity of early societies. Cliodynamics 1, 58–80.

Gronenborn, D., Strien, H.C., Dietrich, S., Sirocko, F., 2014. “Adaptive cycles” and climate fluctuations: a case study from linear pottery culture in western Central Europe. J. Archaeol. Sci. 51, 73–83. http://dx.doi.org/10.1016/j.jas.2013.03.015.

Hodell, D.A., Brenner, M., Curtis, J.H., 2005. Terminal classic drought in the northern Maya lowlands inferred from multiple sediment cores in Lake Chichancanab (Mexico). Quat. Sci. Rev. 24, 1413–1427. http://dx.doi.org/10.1016/j.quascirev.2004.10.013.

Hu, F.S., 2003. Cyclic variation and solar forcing of Holocene climate in the Alaskan subarctic. Science 301, 1890–1893. http://dx.doi.org/10.1126/science.1088568.

Koopmans, L., 1974. The Spectral Analysis of Time Series, 2nd ed. Academic Press, New York.

Langdon, P.G., Barber, K.E., Hughes, P.D.M., 2003. A 7500-year peat-based palaeoclimatic reconstruction and evidence for an 1100-year cyclicity in bog surface wetness from Temple Hill Moss, Pentland Hills, southeast Scotland. Quat. Sci. Rev. 22, 259–274. http://dx.doi.org/10.1016/S0277-3791(02)00093-8.

Lomb, N.R., 1976. Least-squares frequency-analysis of unequally spaced data. Astrophys. Space Sci. 39, 447–462. http://dx.doi.org/10.1007/bf00648343.

Marcus, J., 1992. Political fluctuations in Mesoamerica: dynamic cycles of Mesoamerican states. Res. Explor. 8, 392–411.

Martinez, M., Kotov, S., De Vleeschouwer, D., Pas, D., Pälike, H., 2016. Testing the impact of stratigraphic uncertainty on spectral analyses of sedimentary series. Clim. Past 12, 1765–1783. http://dx.doi.org/10.5194/cp-12-1765-2016.

Masson, M.A., 2012. Maya collapse cycles. Proc. Natl. Acad. Sci. 109, 18237–18238. http://dx.doi.org/10.1073/pnas.1213638109.

deMenocal, P., 2000. Coherent high- and low-latitude climate variability during the Holocene warm period. Science 288, 2198–2202. http://dx.doi.org/10.1126/science.288.5474.2198.

Moreno, A., Cacho, I., Canals, M., Grimalt, J.O., Sánchez-Goñi, M.F., Shackleton, N.,

Sierro, F.J., 2005. Links between marine and atmospheric processes oscillating on a millennial time-scale. A multi-proxy study of the last 50,000 yr from the Alboran Sea (Western Mediterranean Sea). Quat. Sci. Rev. 24, 1623–1636. http://dx.doi.org/10.1016/j.quascirev.2004.06.018.
Mudelsee, M., 2014. Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, 2nd ed. Springer, Cham, Switzerland.
R Core Team, 2017. R: A Language and Environment for Statistical Computing.
Redman, C.L., Kinzig, A.P., 2003. Resilience of past landscapes: resilience theory, society, and the longue durée. Ecol. Soc. 7.
Rosen, A.M., Rivera-Collazo, I., 2012. Climate change, adaptive cycles, and the persistence of foraging economies during the late Pleistocene/Holocene transition in the Levant. Proc. Natl. Acad. Sci. 109, 3640–3645. http://dx.doi.org/10.1073/pnas.1113931109.
Sandoval, Z.V., 1998. Dynamic Cycles of Mesoamerican States: Perspectives from Group 9J-5 and the Copán Valley. Harvard University, Cambridge, MA.
Schulz, M., Mudelsee, M., 2002. REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. Comput. Geosci. 28, 421–426. http://dx.doi.org/10.1016/S0098-3004(01)00044-9.
Sorrel, P., Debret, M., Billeaud, I., Jaccard, S.L., McManus, J.F., Tessier, B., 2012. Persistent non-solar forcing of Holocene storm dynamics in coastal sedimentary archives. Nat. Geosci. 5, 892–896. http://dx.doi.org/10.1038/ngeo1619.
Telford, R., Heegaard, E., Birks, H., 2004. All age-depth models are wrong: but how badly? Quat. Sci. Rev. 23, 1–5. http://dx.doi.org/10.1016/j.quascirev.2003.11.003.
Thompson, V.D., Turck, J.A., 2009. Adaptive cycles of coastal hunter-gatherers. Am. Antiq. 74, 255–278. http://dx.doi.org/10.2307/20622426.
Turchin, P., Nefedov, S.A., 2009. Secular Cycles. Princeton University Press, Princeton, N.J. http://dx.doi.org/10.1017/CBO9781107415324.004.
Ur, J.A., 2010. Cycles of civilization in Northern Mesopotamia, 4400-2000 BC. J. Archaeol. Res. 18, 387–431. http://dx.doi.org/10.1007/s10814-010-9041-y.
Vaníček, P., 1971. Further development and properties of the spectral analysis by least-squares. Astrophys. Space Sci. 12, 10–33. http://dx.doi.org/10.1007/BF00656134.
Zhang, D.D., Jim, C.Y., Lin, G.C.S., He, Y.Q., Wang, J.J., Lee, H.F., 2006. Climatic change, wars and dynastic cycles in China over the last millennium. Clim. Chang. 76, 459–477. http://dx.doi.org/10.1007/s10584-005-9024-z.
Zimmermann, A., 2012. Cultural cycles in Central Europe during the Holocene. Quat. Int. 274, 251–258. http://dx.doi.org/10.1016/j.quaint.2012.05.014.